EP0778575

Publication Title:

Dynamic random access memory

Abstract:

An inexpensive single-chip integrated DRAM memory system having a high density and large bandwidth is provided comprising a DRAM array 10 having a plurality of pipelined stages 12, a control logic circuit 11 for controlling said DRAM array 10, and buffer means 13 integrated onto said chip for storing data being fetched from said DRAM array. The DRAM array and said control logic and said buffer means are all integrated onto one and the same substrate 1, wherein said control logic 11 generates a control signal for controlling operations taking place in said plurality of pipelined stages and the final stage of said pipeline 12 inputs/outputs data from said buffer means 13 in a burst mode.

Data supplied from the esp@cenet database - http://ep.espacenet.com

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 778 575 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
11.06.1997 Bulletin 1997/24

(51) Int Cl.[6]: **G11C 7/00**, G11C 11/409

(21) Application number: **96308132.8**

(22) Date of filing: **11.11.1996**

(84) Designated Contracting States:
DE FR GB IE

(30) Priority: **06.12.1995 JP 317926/95**

(71) Applicant: **International Business Machines Corporation**
Armonk, N.Y. 10504 (US)

(72) Inventor: **Katayama, Yasunao**
Sagamihara-shi, Kanagawa-ken (JP)

(74) Representative: **Davies, Simon Robert**
I B M
UK Intellectual Property Department
Hursley Park
Winchester, Hampshire SO21 2JN (GB)

(54) Dynamic random access memory

(57) An inexpensive single-chip integrated DRAM memory system having a high density and large bandwidth is provided comprising a DRAM array 10 having a plurality of pipelined stages 12, a control logic circuit 11 for controlling said DRAM array 10, and buffer means 13 integrated onto said chip for storing data being fetched from said DRAM array. The DRAM array and said control logic and said buffer means are all integrated onto one and the same substrate 1, wherein said control logic 11 generates a control signal for controlling operations taking place in said plurality of pipelined stages and the final stage of said pipeline 12 inputs/outputs data from said buffer means 13 in a burst mode.

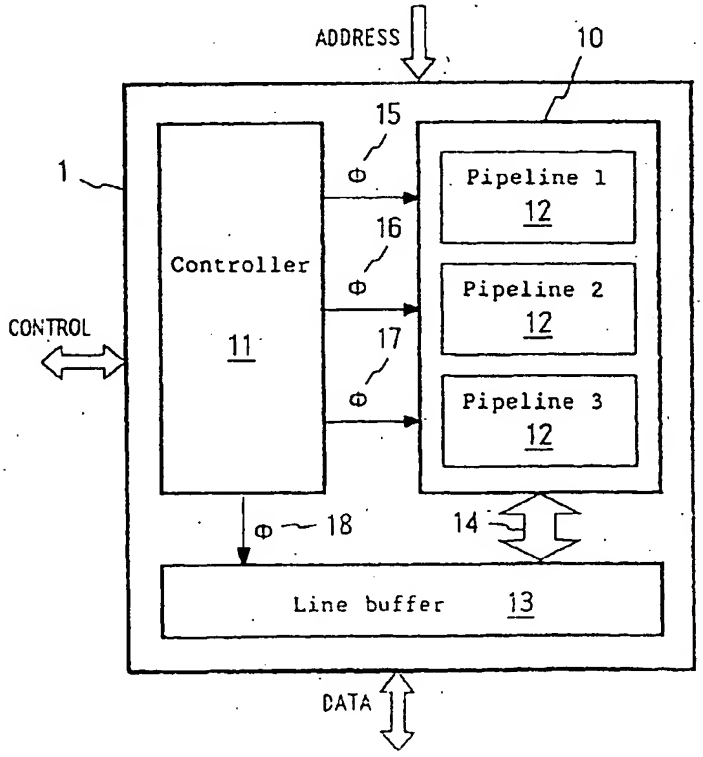

FIG. 4

EP 0 778 575 A2

Description

The present invention relates to a dynamic random access memory (DRAM) as typically used in electronic computer memory systems.

In systems using ultra-high-density DRAMs, it is expected that the whole memory system will be integrated onto a single or a few (on the order of two or three) chips. This is because, in relatively small-sized computers such as personal computers, increases in density resulting from advances in DRAM integration techniques are ahead of increases in the capacity of main memory. For example, a main memory capacity of around 8 MB is needed for a personal computer but 64 MB of DRAM can be implemented with a single chip. In addition, there is a high possibility of 256 MB DRAMS being put to practical use in the near future.

Under such circumstances, rather than incorporating a DRAM interface into the CPU, integrating the DRAM control logic together with the DRAM array and connecting the integrated chip directly to a CPU bus would improve cost performance. When a great number of DRAM chips are needed to configure a main memory, a scheme for integrating DRAM control logic together with the DRAM array onto a single chip is extremely expensive, since a large integrated area is used for DRAM control logic. The reason for this is that the control logic must be integrated onto each DRAM chip and consequently the total number of DRAM control logic circuits in the memory system increases. However, if the main memory can be composed of a single DRAM chip as is mentioned above, it is only necessary to integrate a single DRAM control logic onto one and the same chip, so that the increase in the chip area is not as significant. In brief, at present, the scheme for incorporating a DRAM interface into the CPU is employed because the scheme for integrating DRAM control logic together with a great number of DRAM arrays leads to an increase in chip size and package cost, and moreover complicates the testing of products. However, as the number of DRAMs used in relatively small-sized personal computers decreases as a consequence of the higher density in recent DRAM integration, the cost problem is being solved. What is also of importance is that employing the scheme for integrating the respective DRAM control logic together with DRAM arrays into a single chip makes it possible to bring about a great advantage in the performance of memories from the following two points of view.

Firstly, in techniques such as synchronous DRAM and burst EDO (Extended Data Out) for the interface of conventional DRAMs, stress is laid on speeding-up the clock. Accordingly, the scheme of incorporating a DRAM interface into the CPU cannot always be satisfactory in making use of bus cycles without waste.

Secondly, optimization of a critical path is more easily achieved between the DRAM array and DRAM control logic than between the CPU and the DRAM control logic. In other words, it requires longer time and more elaboration to optimize a critical path between the CPU and the DRAM control logic. For example, overhead due to the multiplexing of addresses cannot be avoided. Moreover, the speedier operation is, the more difficult it becomes to control skew in clock signals transferred between the chips.

Conventionally, a scheme for the provision of an external cache has been employed to upgrade the data transfer rate of DRAMs. However, the bus utilization ratio of DRAM has improved to equal the bus utilization ratio of SRAM in a cache memory, and moreover if the lead-off cycle (period of time from the initiation of RAS to the initiation of CAS) is shortened by integrating the DRAM control logic together with the DRAM array, a transfer rate comparable to that obtained when an external cache is provided can be implemented. Thus, the merit of attaching an off-chip external cache memory is reduced, or disappears altogether. What is worse, since connection of an external cache necessarily generates an overhead due to the data transfer between the external cache and the DRAM, performance may actually be lowered by inclusion of the cache. This deterioration of performance becomes conspicuous in a multimedia-type of application in which very great quantities of data must be transferred at high speed and therefore a solution using attachment of an external cache has the danger of worsening performance in addition to the increase in cost originating from the external cache.

A critical problem in the operational performance of DRAM in the conventional approach is the low speed of the data transfer rate in the path of row access. Speeding-up of the data transfer rate in column access has been fully studied e.g., in techniques such as synchronous DRAMs. However, research has not made as much advance in the data transfer rate for row access as in the data transfer rate for column access. That is, with the increasing data transfer rate in column access, the relatively slower row access is becoming the critical path in 4-beat burst mode operation. In fact, since the page miss ratio between consecutive burst actions (probability of necessary data being present in the same row in the next access) can be as much as 50%, row access takes place fairly frequently.

As a solution to this problem, it is possible to raise the data transfer rate by an appropriate pipelined operation in the interface between the CPU-DRAM or in the row access process for the DRAM (e.g., the address pipeline of the CPU). In a DRAM, individual steps of sensing, writing back and precharging must be accomplished in that order. Accordingly, the array time constant (the total time taken for sensing, writing back and precharging) becomes the time required for the data transfer of the DRAM. In data transfer between DRAMs, not only the time for the operational steps of sensing, writing back and precharging, but also the time for selecting row addresses and column addresses is necessary. Pipelining enables the time taken for selection of

these addresses to be concealed behind the operational steps of sensing, writing back and precharging. Thus, the period of time for accessing two successive row addresses (the time taken for selection of the row address + the time taken for sensing, writing back and precharging, referred to as the RAS cycle hereinafter) can be made close to the time taken for sensing, writing back and precharging (referred to as the array time constant hereinafter). Certain conventional techniques disclose pipelined DRAMs having a row access path. In none of these conventional techniques, however, does the RAS cycle reach the array time constant.

The points mentioned above, especially the high integration of DRAM chips owing to the advance in LSI technique and the flow of the background art will be outlined referring to Figures 1-3. Figure 1 shows an ordinary memory system using DRAM. This memory system is divided into numerous DRAM chips 10. This is because the integrated density of a current DRAM chip is too small for the capacity required for uses of the main memory so that a single DRAM chip cannot constitute the main memory. To control the plurality of DRAM chips, an off-chip DRAM controller 11 is required between the CPU (not shown) and the DRAM chips 10. As shown in Figure 1, a DRAM controller 11 receives a memory access request from the CPU and supplies a row address and column address. Data outputted from a DRAM chip 10 is transferred through the DRAM controller 11 to the CPU. Data inputted to a DRAM chip 10 is processed similarly. However, according to such an off-chip scheme, the signal path between the controller 11 and a DRAM chip 10 lengthens and accordingly it is difficult to synchronize the addressing with the other operations of the DRAM with control of the delay of the control signal for RAS, CAS and the like. This difficulty becomes conspicuous especially in the case of high-speed data transfer.

Advances in the high-density integration techniques for DRAM enable the memory capacity required for a relatively small-sized computer to be almost satisfied with one or at most a few DRAM chips. As a result, the DRAM array 10 (corresponding to the DRAM chips in Figure 1) can be integrated onto the same chip 1 as that of the DRAM controller 11, as shown in Figure 2. In other respects, Figure 2 is similar to Figure 1. The technique relevant to Figure 2, in which both are integrated on one and the same chip 1, is expected to upgrade operational performance and cost-saving somewhat, as compared with that of Figure 1, in that one level of packaging is omitted. However, the upgrade of performance does not differ greatly from that obtained when the DRAM controller 11 is incorporated into the CPU and consequently has no technically significant effect.

Figure 3 shows a DRAM configuration using pipelining of synchronous DRAMs or the like. A DRAM chip 10 is characterized by having a DRAM pipeline 12 formed inside it. However, this DRAM pipeline 12 is controlled by an external RAS (row address strobe) and CAS (column address strobe). Thus, the control of pipelines inside individual DRAMs becomes restrictive.

Several techniques other than pipelining have so far been proposed to upgrade the operational speed of DRAM.

- Page hit scheme

The page hit scheme is a scheme to utilize the sense amplifier in a DRAM as a cache as well. The locality of data or an instruction in a page (here designating a row in the DRAM) is effectively utilized under a normal page mode in the DRAM. However, when an on-chip cache is connected to the CPU, the page hit ratio between two continuous cache line transfers is not so high, and the larger the capacity of the on-chip cache is, the lower the page hit ratio becomes. Thus, considering the delay for comparison of tags at the time of a page miss, the precharging time and the like, a significant upgrade in the operational speed of a DRAM cannot be expected even if the page hit scheme is employed.

- Interleave scheme

The interleave scheme is a scheme for dividing a memory into a plurality of banks and accessing these banks in sequence. This scheme is also often employed to raise the data transfer rate of DRAM. In the case of off-chip interleaving for a relatively small-sized computer, however, the granularity of memory (designating the minimum unit of additive memory installation) increases. For example, when a memory is divided into two banks, granularity doubles, which is a serious problem in an ultra-high-density DRAM. On the other hand, in the case of on-chip interleaving, the bandwidth can be increased without a great effect on the granularity of the memory. However, generally speaking, this technique is effective in upgrading the operational speed of DRAM only at the time of operations where different memory banks are alternately accessed. Thus, it lacks generality as a scheme for upgrading the data transfer rate.

Accordingly, the present invention provides a dynamic random access memory (DRAM) system comprising a chip, a DRAM array mounted on said chip having a plurality of pipelined stages, and control logic mounted on said chip for controlling said DRAM array, said control logic including means for generating a signal for controlling said plurality of pipelined stages.

In the preferred embodiment, the chip further comprises buffer means for storing data being fetched from said DRAM array, and to input/output said data in a burst mode. A plurality of pipelined stages comprise a first stage for setting an address from which data is to be fetched and a second stage for performing internal memory operations in said DRAM array. Said control logic is connected to a Central Processing Unit (CPU) which is not mounted on said chip.

The invention also provides a method for operating a dynamic random access memory (DRAM) system, said DRAM system having a chip, a DRAM array mounted on said chip having a plurality of pipelined stages, control logic mounted on said chip for controlling said DRAM array, and buffer means for storing data being fetched from said DRAM array, said method including the steps of:

generating a signal at the control logic for controlling said plurality of pipelined stages; and
inputting/outputting said data in said buffer means in a burst mode;
wherein a first period to input/output said data in a burst mode is longer than a second period for performing an internal memory operation in said DRAM array.

In the preferred embodiment, there are substantially no clock cycles between a first data output and a second data output, both in a burst mode.

Such an approach provides a high-density and low-cost DRAM memory array having a high data transfer rate. The DRAM control logic is integrated together with a pipelined DRAM array onto a single chip and optimizes access to the DRAM array, thereby reducing cost and improving data transfer rate. In addition, the leadoff and precharge cycle in DRAM access is less than in the prior art, this minimization being possible for both write and readout access and for both single-beat access and continuous-beat access (burst mode). The above approach further provides a memory system comprising a single chip or a few chips without the need for addition of a cache memory using SRAM cells and without lowering operational performance. A memory system comprising a single chip or a few chips is particularly applicable to multimedia use in which a large quantity of data is transferred at high speed.

In the preferred embodiment, a single-chip DRAM system comprises a DRAM array with a plurality of pipelined stages, control logic for controlling said DRAM array, and buffer means for storing data fetched from said DRAM array, all of which are integrated onto one and the same chip. In this DRAM system, the control logic generates a signal for controlling operations taking place in said plurality of pipelined stages. In addition, the final stage in the plurality of stages is a stage for inputting/outputting data from said buffer means in burst mode. Consequently, for some burst lengths the period of time for inputting/outputting data in burst mode can be made longer than that for internal operation in the DRAM array, and therefore the internal operation in the DRAM array can be perfectly hidden behind data inputting/outputting operation.

It is important to note that the DRAM control logic and the DRAM array are mounted together on a single chip, and that the pipelined DRAM is controlled with a clock generated by the on-chip DRAM control logic. Further, synchronous burst transfer is executed by a line buffer as the final stage in the pipeline.

A single-chip integrated main memory is typically constructed for the purpose of excluding an external cache memory, but can be referred to as a fully-associative DRAM cache memory with a 100% hit ratio from another point of view. It can be said that the cache data memory is enlarged to a size close to that of the main memory and accordingly the step of comparing tags becomes unnecessary. Such features are most suitable for multimedia use. Multimedia use requires high-speed transfer of an extremely large quantity of data and a conventional cache scheme is not applicable to multimedia. In other words, a conventional cache DRAM does not have sufficient performance to meet this requirement. This is because the cache hit ratio is restricted by the size of the SRAM and the data transfer rate is restricted by the bandwidth of the DRAM.

These points are summarized as follows:

(1) The preferred embodiment has a DRAM control logic and a DRAM array mounted on one and the same chip. This is because use of an ultra-high-density DRAM enables the main memory to consist of a single or a few DRAM chips.

(2) The preferred embodiment controls a pipelined DRAM by means of a clock generated by a DRAM control logic. If joined to (1), this has a great significance. That is, since no clock skew between the external control logic and the DRAM array should develop, unlike in the background art, extremely rapid data transfer operations can be implemented.

(3) The preferred embodiment executes a synchronous burst transfer from a line buffer as the final stage of the pipeline. Since the final stage becomes a stage corresponding to an address decoding operation and array operation of a DRAM, it becomes possible to output data to or input data from the outside in coordination with these operations. Thus, the bus cycle becomes fully usable.

(4) The above features enable the period for data input/output in burst mode to be made longer than that for operations inside the DRAM array. Thus, optimum data transfer operation can be achieved without interposing a clock cycle between the burst transfers.

A preferred embodiment of the invention will now be described in detail by way of example only with reference to the following drawings:

Figure 1 shows a memory system using a DRAM according to the prior art;
Figure 2 shows another memory system using a DRAM according to the prior art;

Figure 3 shows still another memory system using a DRAM according to the prior art;
Figure 4 is a general view of a DRAM system in accordance with the present invention;
Figure 5 is a drawing showing a DRAM system according to the present invention;
Figure 6 is an operation chart for a single-beat mode; and
Figure 7 is an operation chart for a 4- beat burst mode.

Referring to the drawings, preferred embodiments of the present invention will be disclosed hereinafter. First, the difference between the background art, as seen from the standpoint of the flow of technology seen in Figure 1 to Figure 3, and a preferred embodiment of the present invention will be described from the view point of the DRAM configuration and control scheme by using Figure 4.

A preferred embodiment of the present invention is shown in Figure 4, in which a plurality of pipelines 12 are used in a DRAM array 10 and the respective pipelines are controlled by a DRAM controller 11 which are all integrated onto a single chip 1. In addition, a line buffer 13 is integrated onto the same chip 1. This line buffer 13 can be understood as the final stage in the pipeline. The DRAM array 10 and the line buffer 13 are connected by a wide-bandwidth on-chip bus 14. Simultaneously with operation of each pipeline, data can be transferred between the DRAM array 10 and the line buffer 13. The DRAM controller 11 is also integrated on-chip onto the chip 1. The DRAM controller 11 controls bidirectional data transfer between each stage in the pipeline 12 and the DRAM array 10/line buffer 13 with independent clock signals 15, 16, 17 and 18. According to this scheme, more flexible and more precise control related to the pipeline 12 becomes possible. This is because a limit need not be placed on the number of control signals for the DRAM and only a small skew is present between the clock signals 15-18 used for control. To raise the data transfer rate, read/write has conventionally been performed in a plurality of line buffers, but here a single line buffer is used for the whole system.

Figure 5 shows an example in which a pipelined DRAM is accessed through a DRAM controller 11. Here, the pipeline comprises three stages 20, 30 and 40. These stages include the final stage, a line buffer stage 40 for data transfer. The controller 11 generates three clock signals 1, 2 and 3 for control. The first clock signal 1 controls operations related to addresses for the pipelined DRAM. Thus, this signal corresponds to a RAS clock or CAS clock in the background art. Please note that since in the preferred embodiment, no address multiplexing is used, row and column addresses are fetched at the same time. The second clock signal 2 controls the operational processes in the memory array. The third clock signal 3 controls the processes in data transfer by being supplied to the line buffer 13.

Hereinafter, a memory access operation of the DRAM will be divided into the following cycles, and simplified for purposes of description:

- RD: Row address decode path (path directly before the word line driver, relating to input and decoding of row addresses);
- CD: Column address decode path (path directly before the bit switch, relating to input and decoding of column addresses);
- SE: Path for sensing the potential of a cell in the array;
- WB: Path related to operations for writing back the potential sensed in a cell in the array;
- PR: Path related to precharging operations of the array for equating the bit line potential to the intermediate potential; and
- TR: Path transferring data from the line buffer to an external bus.

When the individual paths are classified/defined in such a manner, the following three parameters may be issues in the design of an interface for DRAM:

- RAS cycle (RD + SE + WB + PR);
- Array time constant (SE + WB + PR); and
- Transfer cycle (product of the number of continuous bits relevant to a burst transfer and the TR).

Referring to Figure 5, the first signal 1, the second signal 2 and the third signal 3 are timing signals to be supplied to means 20 in charge of decoding, means 30 in charge of operation inside the array and means 40 in charge of data transfer, respectively. Incidentally, these means are schematic, and do not necessarily indicate corresponding real structure or precise mechanisms. As shown in Figure 5, the decoding of a row address (RD) and the decoding of a column address (CD) take place in means 20 in charge of decoding. As a result of RD, a predetermined word line (WL) is chosen, whereas a predetermined bit switch (Bit-SW) is chosen as a result of CD. The timing for these is controlled by the first signal 1 supplied by the controller 11. For a predetermined word line that is accessed, means 30 in charge of operation in the array executes sensing (SE), writing back (WB) and precharging (PR). The timing for these is controlled by the second signal 2 supplied by the controller 11. After the completion of this, a bit coupled to a predetermined bit switch from said predetermined word line that was sensed is chosen by making use of the result of CD. The timing for this is also controlled by the second

signal 2. While being controlled by the second signal 2, the data are transferred through the internal bus 14 to the line buffer 13 (means 40 in charge of data transfer). Thereafter, while being controlled by the control signal 3 supplied from the controller 11, data are transferred to the external bus. Transfer can employ either a single mode or a burst mode.

One feature is that, together with achieving the pipelining of operation by the provision of means 20 in charge of decoding (referred to as decode handling means 20 hereinafter), means 30 in charge of operation inside the array (referred to as internal operation handling means 30 hereinafter) and means 40 in charge of data transfer (referred to as data transfer means 40 hereinafter), individual pipelines are made controllable with the first signal 1, the second signal 2 and the third signal 3 independently supplied from the controller 11, respectively.

Another feature is that a controller 11 integrated onto the same chip as the DRAM array 10 controls individual pipeline means independently by generating a plurality of signals 1, 2 and 3, which enables the timing signals to be controlled with higher accuracy than achieved by the control of RAS and CAS in the background art with the DRAM control logic installed outside the chip on which the DRAM is integrated. As a result, an effect of drastically speeding the fetching of data from the DRAM array is obtained.

Furthermore, the line buffer 13 is integrated onto the same chip as the DRAM array 10, which enables data to be burst-transferred at a very high frequency. As with the speeding up of data fetching from the DRAM array, this is because the pipelined means relevant to burst transfer is controlled with a signal generated by the controller 11 integrated onto the same chip as the DRAM array 10. In addition, because the DRAM array 10 and the line buffer 13 are connected through an internal bus 14 that is large in bandwidth, data fetching can follow the speeding-up of the data transfer rate.

The reason for the generating of these advantages will be described assuming that one clock cycle is necessary for each cycle. Figure 6 shows a timing chart for operation of the pipelined DRAM array in a single-beat transfer mode in the preferred embodiment.

Referring to Figure 5, the restrictive conditions (rules) for operation of a DRAM system in the preferred embodiment are clarified as follows:

(1) In decode handling means 20, internal operation handling means 30, and data transfer means 40, the respective necessary operations are carried out independently. Accordingly, operation in each means is operable at mutually independent timing without being influenced by the timing of operations in other means. This is a principle in operation of such a DRAM system. For example, RD/CD is performed in decode handling means 20 and PR is performed in internal operation handling means 30 independently of each other. Accordingly, RD/CD and PR are simultaneously operable in the same clock cycle.

(2) Internal operation handling means 30 never fails to operate in the particular sequence of SE, WB and PR, and SE of the next cycle can start only after PR of the preceding cycle terminates.

(3) In a read mode, though RD and SE are performed in different means, SE cannot operate unless an address is supplied and therefore RD in decode handling means 20 must always precede SE (exception of (1)).

As compared with a conventionally known access method for a DRAM, this approach has the following advantages in the aspect of performance. First, since a DRAM controller is integrated onto the same chip as the DRAM cell, no multiplexing of addresses (alternate supply of row addresses and column addresses from the same pin) is necessary. Consequently, an address relevant to a row and column is immediately supplied and the decoding thereof can start simultaneously when an address and a start signal are supplied by the CPU. Secondly, internal pipelining of the DRAM enables the RD/CD stage after the start of operation to be concealed by the preceding array operation (SE + WB + PR). On referring to Figure 6 in this respect, the cycle 90 extending from one RD to the next RD (hereinafter, referred to as the RAS time constant) has changed from 4 cycles to 3 cycles. Accordingly, data transfer TR also becomes possible every three cycles. For example, when read modes continue, a first TR 100 and a second TR 101 can arrive every 3 cycles. This is because RD becomes possible simultaneously with PR from the second access operation (e.g., PR 107 and RD 108 take place simultaneously).

Next, in switching from read mode to write mode, since TR becomes possible simultaneously with SE in write mode, TR can take place at intervals of only one clock cycle. For example, after TR 101 was accomplished in read mode, execution of TR 102 in write mode requires only 1 clock cycle for address decoding RD/CD in write mode between them.

On the contrary, on switching from write mode to read mode, a larger interval appears between TR in write mode and TR in read mode. For example, an interval as long as 3 clock cycles becomes necessary between TR 102 in write mode and TR 103 in read mode. This is because SE and TR can be executed in write mode (cf. 102) and the next SE 106 can start only after the preceding PR 105 terminates (rule 2 above).

As described above, the RAS time constant on switching of write mode/read mode becomes 3 clocks on average. Thus, it is concluded that TR can be executed at intervals of 2 cycles on average or at intervals of 3 cycles regardless of whether there is mode switch-

ing or not.

As shown in Figure 7, such advantages in data transfer become even more conspicuous when continuous 4-beat burst transfer takes place. Since it is normal for the CPU to have a first-order on-chip cache memory, the continuous burst mode is the operation mode that will be most usually performed for the operation of reading DRAM. As shown in Figure 7, the TR stage continues during 4 clock cycles when a continuous 4-beat burst transfer takes place. Accordingly, the array time constant that continues during 3 clock cycles (SE + WB + PR) can be concealed by a TR stage that continues for 4 clock cycles, which provides a bandwidth optimized for a fixed bus cycle length. Even if the bus cycle length is not fixed, uninterrupted data transfer is assured in read mode when the array time constant is shorter than the transfer cycle relevant to a continuous burst. And as long as use is made of a standard TTL interface, to effectively utilize a bus in this manner is highly important since the frequency of a bus clock on a planar board is limited. However, even with such an approach, a time lag of 3 clock cycles cannot be eliminated at the switching from a write mode to a read mode.

Figure 7 will be further described. In this case, the following rule is further added to the rules mentioned above.

(4) When read mode continues, the next TR cannot start unless the preceding TR terminates.

According to this rule, when read mode continues, the next TR 201 cannot start unless the preceding TR 200 terminates. Thus, it is sufficient if the corresponding SE terminates by the start time of the next TR 201. In write mode, WB and PR must take place only after all TRs for write terminate. Accordingly, a large lag time occurs at the next time of switching to read mode. To be brief, an interval as long as 3 clocks is required between the last TR 202 in write mode and the first TR 203 in read mode. This is because the next SE can take place only after the preceding PR (rule 2 above).

However, at least when read mode continues, or for a 4-beat burst mode when read mode switches over to write mode, its continuous bus use is ensured, which is highly effective in the sense of raising the data transfer rate.

The approach described above somewhat resembles the pipeline scheme in the background art, especially the synchronous DRAM in pipelining, and in using burst transfer, the burst EDO scheme in the background art. Furthermore, as a general scheme for raising the data transfer rate in the background art, the cache DRAM is already known. Hereinafter, the superiority of the approach described above will now be pointed out as compared with these conventional techniques.

- Pipeline scheme

A pipeline scheme means a scheme for dividing memory system operation into several operational blocks and actuating the blocks in parallel. The employment of this scheme in a column access bus such as SRAM or synchronous DRAM is already known. With these techniques, however, pipelining is fulfilled in single access mode only.

The preferred embodiment differs from conventional pipelining in that pipelining of paths for accessing rows and pipelining of synchronous burst transfer from the line buffer are united. In short, this appears to be an optimal solution for achieving an upgrade in the data transfer rate of DRAM. In addition, with the background art, the control logic for a DRAM was not integrated together with a DRAM array onto one and the same chip (on-chip integration). By contrast, on-chip integration of the DRAM control logic with a DRAM array in the preferred embodiment enables an extremely speedy data transfer rate to be supported because the DRAM array is controlled by the on-chip integrated DRAM control logic. In other words, except with on-chip integration of the DRAM control logic, control of high-speed pipelined burst transfer becomes extremely complicated and difficult. Thus with increasing clock speed, it becomes extremely difficult on account of clock skew and the like to execute control of high-speed and precise timing by using an off-chip control logic.

- Synchronous DRAM

The synchronous DRAM is a scheme for synchronizing serial read/write operations by using an external clock. The preferred emboidment might conceivably be regarded as a type of synchronous DRAMs but is distinguished by low cost, no need for address multiplexing and possession of a single bank as compared with a conventional synchronous DRAM. Advantages of the preferred embodiment over a conventional synchronous DRAM are as follows:

(1) A conventional synchronous DRAM has required a great deal of labor in testing because of having a great variety of access modes. In contrast, with a DRAM in the preferred embodiment, the types of access modes are limited due to their optimization as uses for main memory and therefore a small amount of labor in testing will suffice.

(2) As a result of a reduced number of access modes, the preferred embodiment can support a burst transfer of the CPU without setting the internal mode register or without interrupting a burst access midway and therefore a small overhead is ensured in memory access.

(3) A conventional synchronous DRAM does not dispense with the multiplexing input of addresses, whereas the preferred embodiment does not need this and accordingly can be satisfied with a shorter leadoff time than that of a conventional synchro-

nous DRAM. In addition, since setting of the mode register is unnecessary, there is another merit of being easier to use.

(4) As compared with a conventional synchronous DRAM, only a small modification from the standard DRAM design will suffice.

(5) Provision of an additive pipelined stage in a path for decoding a row address enables the bandwidth of a memory to approach extremely close to the array time constant.

- Burst EDO scheme

In normal page mode, a burst transfer is also terminated with the termination of a column address strobe. With a burst EDO scheme, however, data continues to be burst-outputted till the start of the next column address strobe. In this respect, the data transfer rate can be increased according to this scheme. The preferred embodiment somewhat resembles the burst EDO scheme in executing a burst transfer. However, the preferred embodiment can further speed up the data transfer rate in comparison to the burst EDO scheme. This is partly because no address decoding for selection of a bank is performed in the memory controller, partly because no multiplexing of addresses is required, and partly because latching of an addresses and transfer of data are synchronized directly with the external clock.

- Cache DRAM

A cache DRAM is a DRAM having an on-chip cache memory connected to the DRAM body through a large-bandwidth internal bus. The preferred embodiment is superior to a cache DRAM in the following respects.

Without use of SRAM or cache control logic, the preferred embodiment can implement a single-chip integrated memory system of equal performance by a combination of the DRAM control logic and the pipelined DRAM array. Thus, lower price and smaller power consumption are provided than for a cache DRAM.

As with the cache DRAM, in the preferred embodiment, the maximum bandwidth of data is also restricted by the external bus clock. The operational performance is determined by the address pipeline ratio for the preferred embodiment and by the cache hit ratio for the cache DRAM. with the preferred embodiment, however, the penalties such as delay due to an address pipeline miss are extremely small. This is because the DRAM control logic is integrated into the same chip as the pipelined DRAM array.

Claims

1. A dynamic random access memory (DRAM) system comprising a chip (1), a DRAM array (10) mounted on said chip having a plurality of pipelined stages (12), and control logic (11) mounted on said chip for controlling said DRAM array, said control logic including means for generating a signal for controlling said plurality of pipelined stages.

2. A DRAM system as recited in claim 1, further comprising buffer means (13) for storing data being fetched from said DRAM array, and to input/output said data in a burst mode.

3. A DRAM system as recited in claim 1 or 2, wherein said plurality of pipelined stages comprise a first stage for setting an address from which data is to be fetched and a second stage for performing internal memory operations in said DRAM array.

4. A DRAM system as recited in any preceding claim, wherein said control logic is connected to a Central Processing Unit (CPU) which is not mounted on said chip.

5. A method for operating a dynamic random access memory (DRAM) system, said DRAM system having a chip (1), a DRAM array (10) mounted on said chip having a plurality of pipelined stages (12), control logic (11) mounted on said chip for controlling said DRAM array, and buffer means (13) for storing data being fetched from said DRAM array, said method including the steps of:

   generating a signal at the control logic for controlling said plurality of pipelined stages; and
   inputting/outputting said data in said buffer means in a burst mode;
   wherein a first period to input/output said data in a burst mode is longer than a second period for performing an internal memory operation in said DRAM array.

6. A method as recited in claim 5, wherein there are substantially no clock cycles between a first data output and a second data output, both in a burst mode.

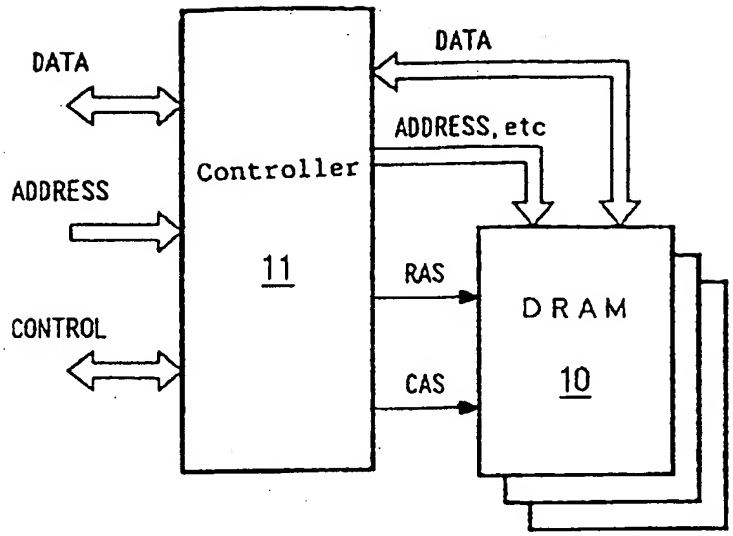
DATA
ADDRESS
CONTROL
Controller
11
DATA
ADDRESS, etc
RAS
CAS
DRAM
10

FIG. 1

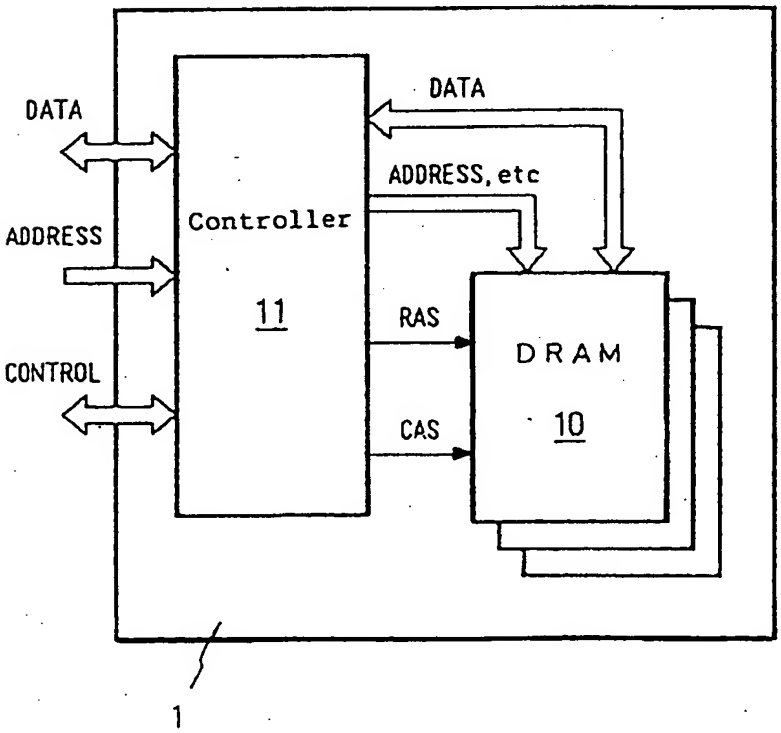
DATA
ADDRESS
CONTROL
Controller
11
DATA
ADDRESS, etc
RAS
CAS
DRAM
10
1

FIG. 2

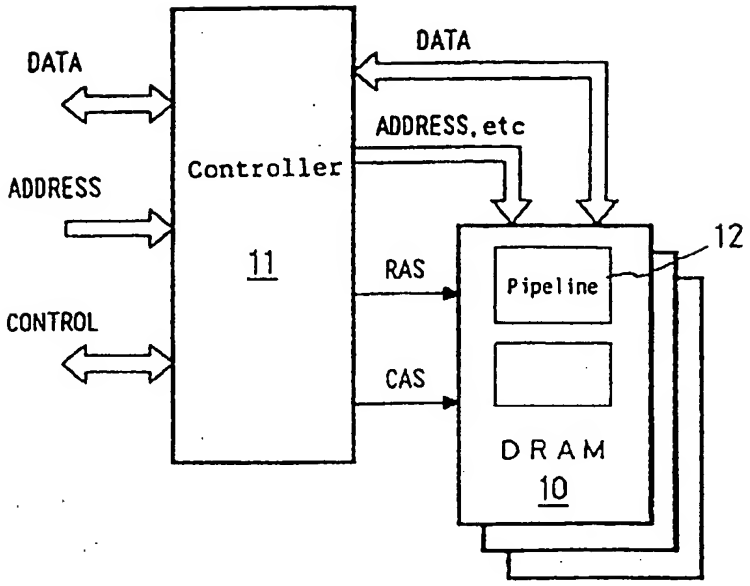
DATA
ADDRESS
CONTROL
Controller
11
DATA
ADDRESS, etc
RAS
CAS
Pipeline
12
DRAM
10

FIG. 3

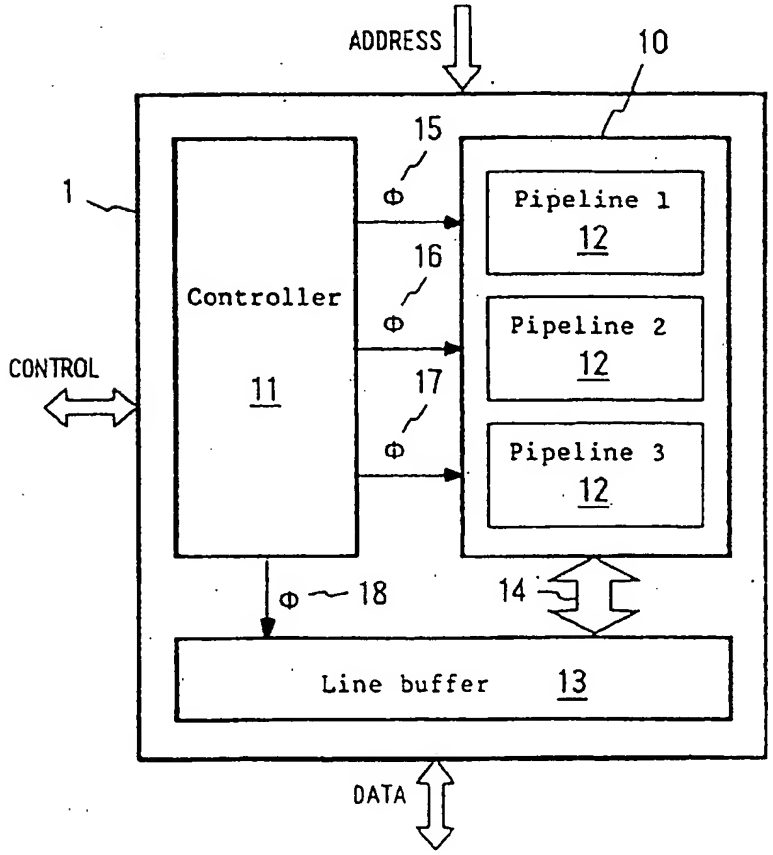
ADDRESS
10
1
15
Φ
16
Φ
17
Φ
Controller
11
CONTROL
Pipeline 1
12
Pipeline 2
12
Pipeline 3
12
Φ ~ 18
14
Line buffer 13
DATA

FIG. 4

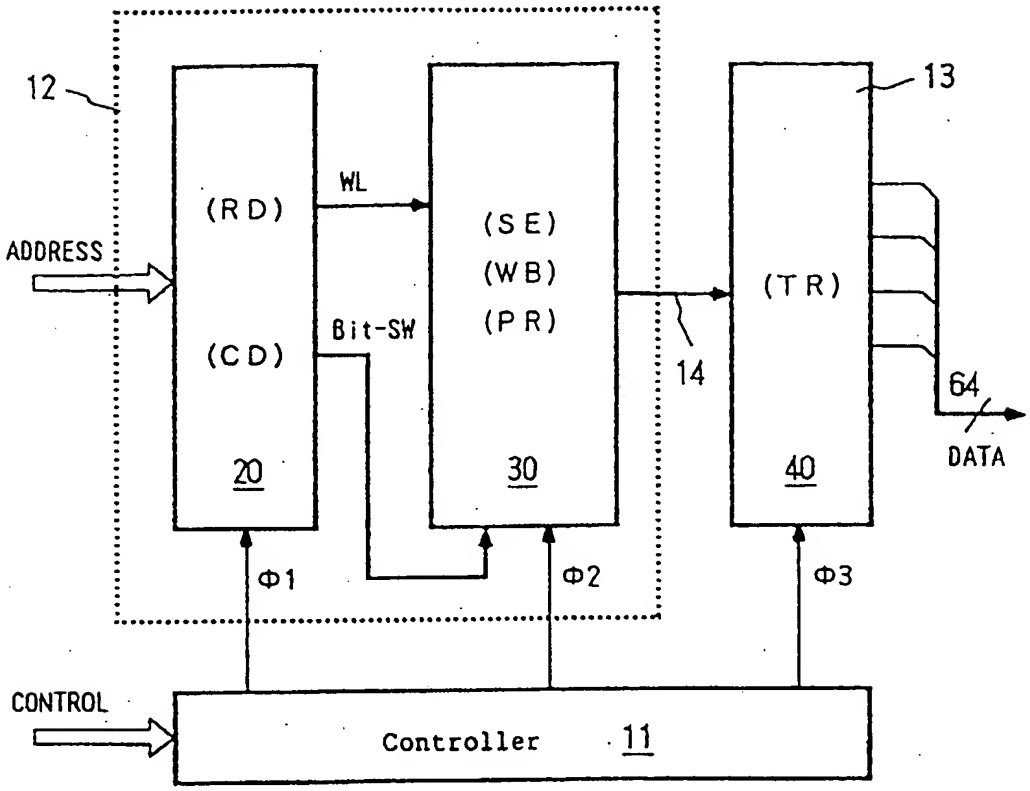
12
13
(RD)
(CD)
20
WL
Bit-SW
(SE)
(WB)
(PR)
30
14
(TR)
40
64
DATA
ADDRESS
Φ1
Φ2
Φ3
CONTROL
Controller 11

FIG. 5

BEST AVAILABLE COPY

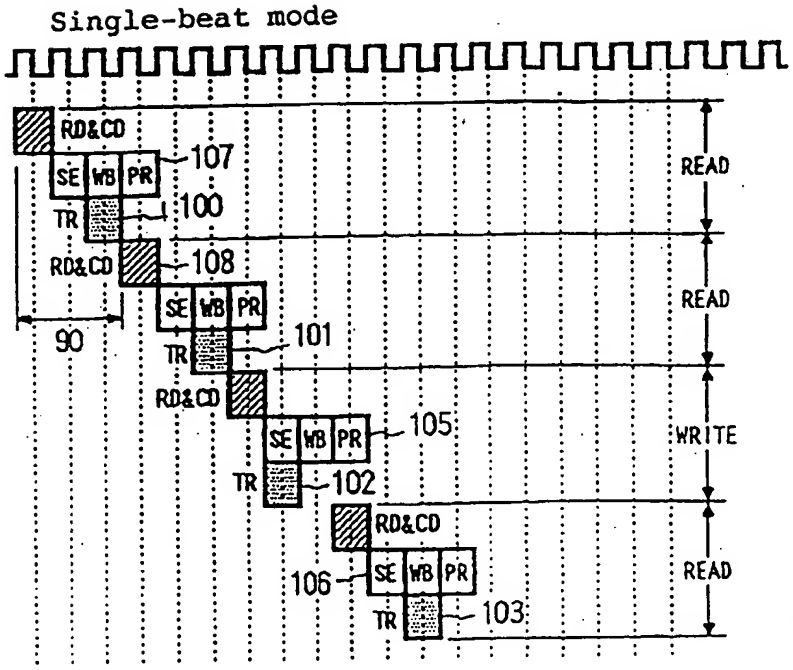

FIG. 6

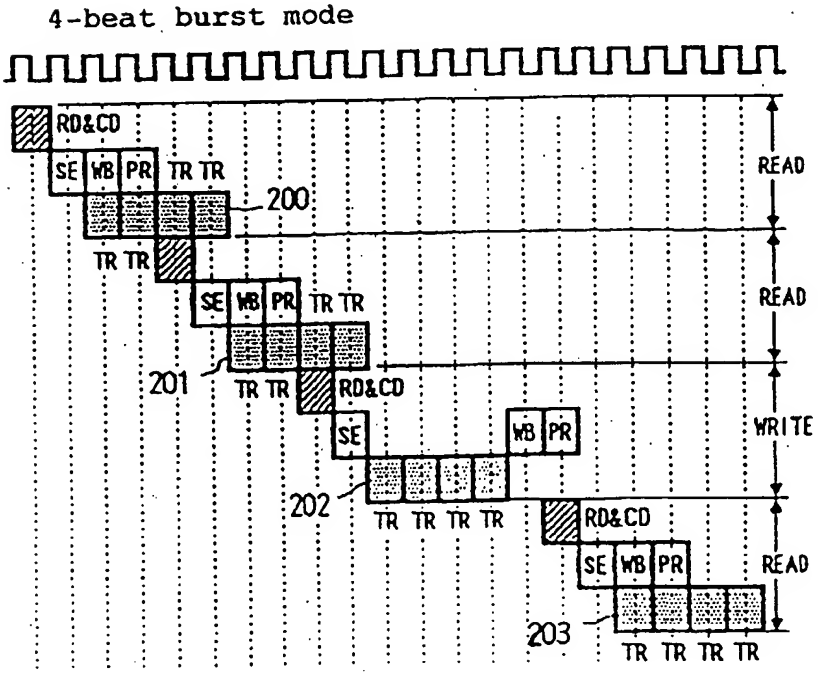

FIG. 7